# NRES_798_7_201501

Statistical methods and linear models

# Basic statistical model

$$Y = deterministic\ part + stochastic\ part$$

Univariate
Multivariate

Linear
Nonlinear
Smoothed

Distribution
Heterogeneity
Auto-correlation
Nested data
(random effects)
Random noise

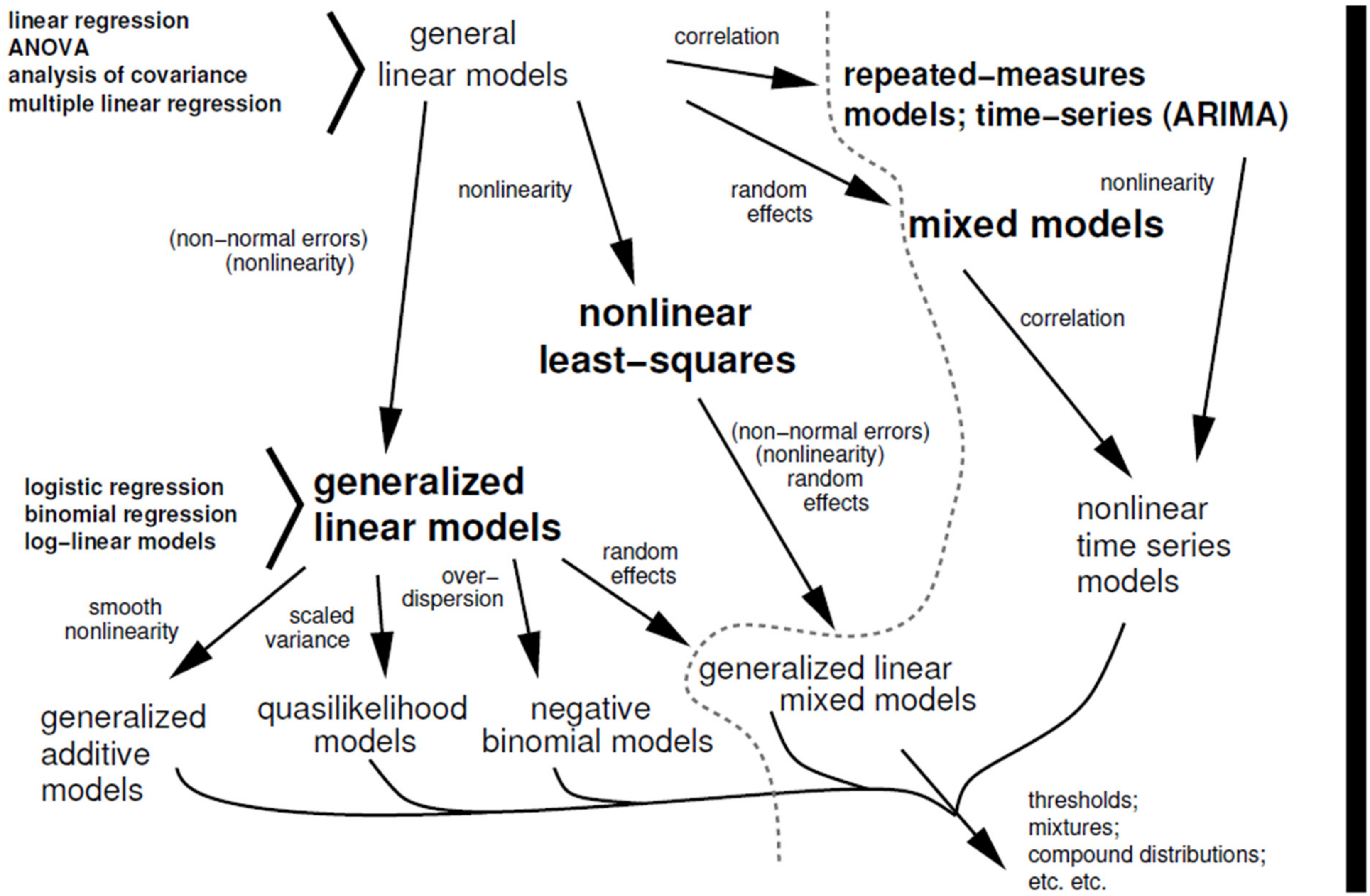
Error term ($\varepsilon_i$)

# General linear models

- Linear regression
- One- and multi-way analysis of variance (ANOVA)
- Analysis of covariance (ANCOVA)

- In R these procedures are use the function lm()

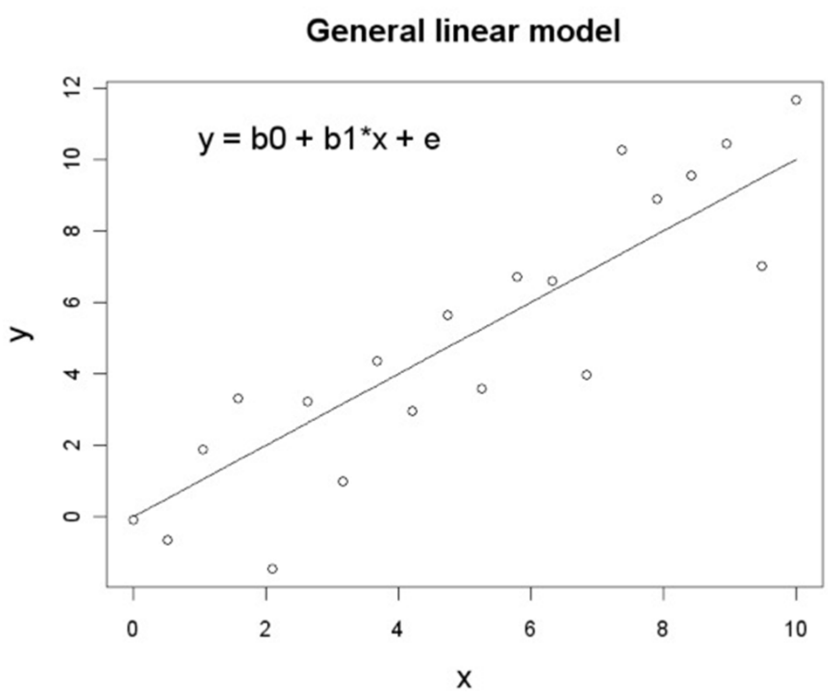
# Beyond linear models




Bolker 2008

# General linear models

$$Y \sim Normal(\overbrace{b_0 + \ b_1 X}^{\text{Deterministic part}}, \underbrace{\sigma^2}_{\text{Stochastic part}})$$




Models that are linear functions of the parameters, not necessarily of the independent variables

# General linear models

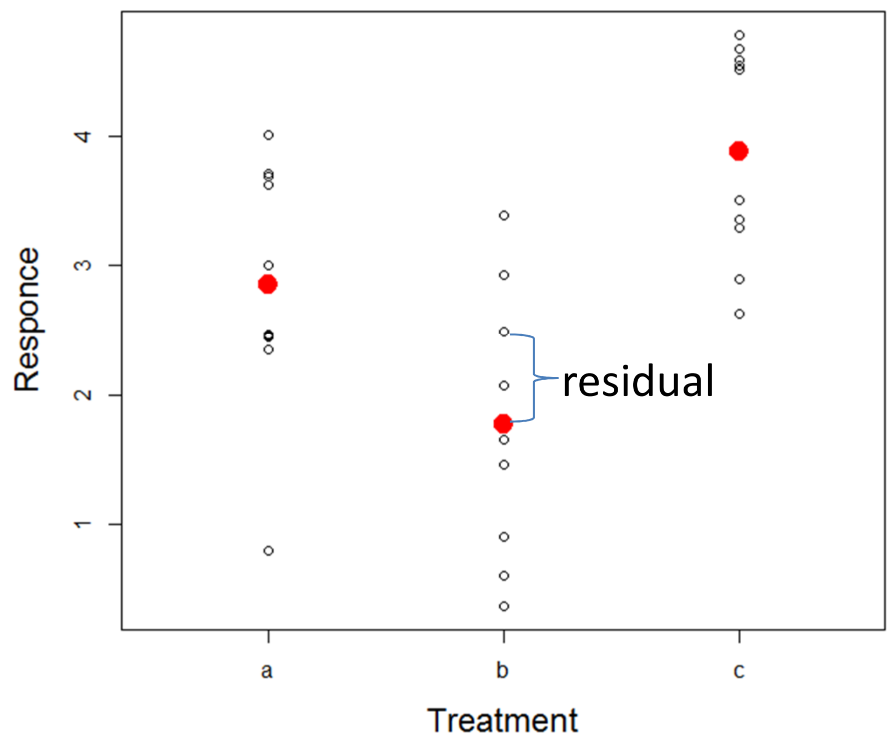
- Assumptions $Y \sim Normal(b_0 + \ b_1 X, \sigma^2)$

  - All observed values are:
    - Independent
    - Any continuous predictor variables (covariates) are measured without error
    - Constant variance (homoscedastic)
    - Normally distributed

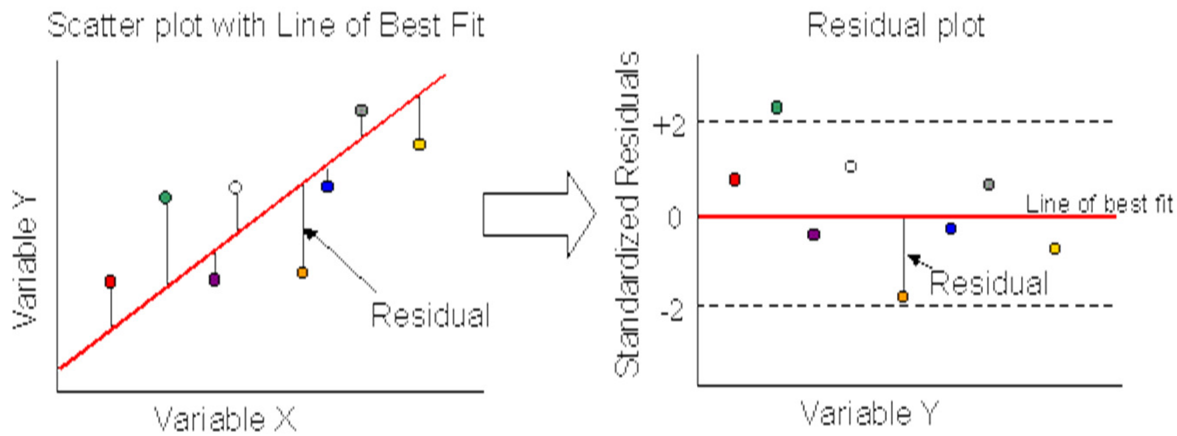
# Homoscedasticity

- Variance of all treatment groups needs to be approximately equal to each other
- Variance needs to be constant across all predictor variables
- Residuals

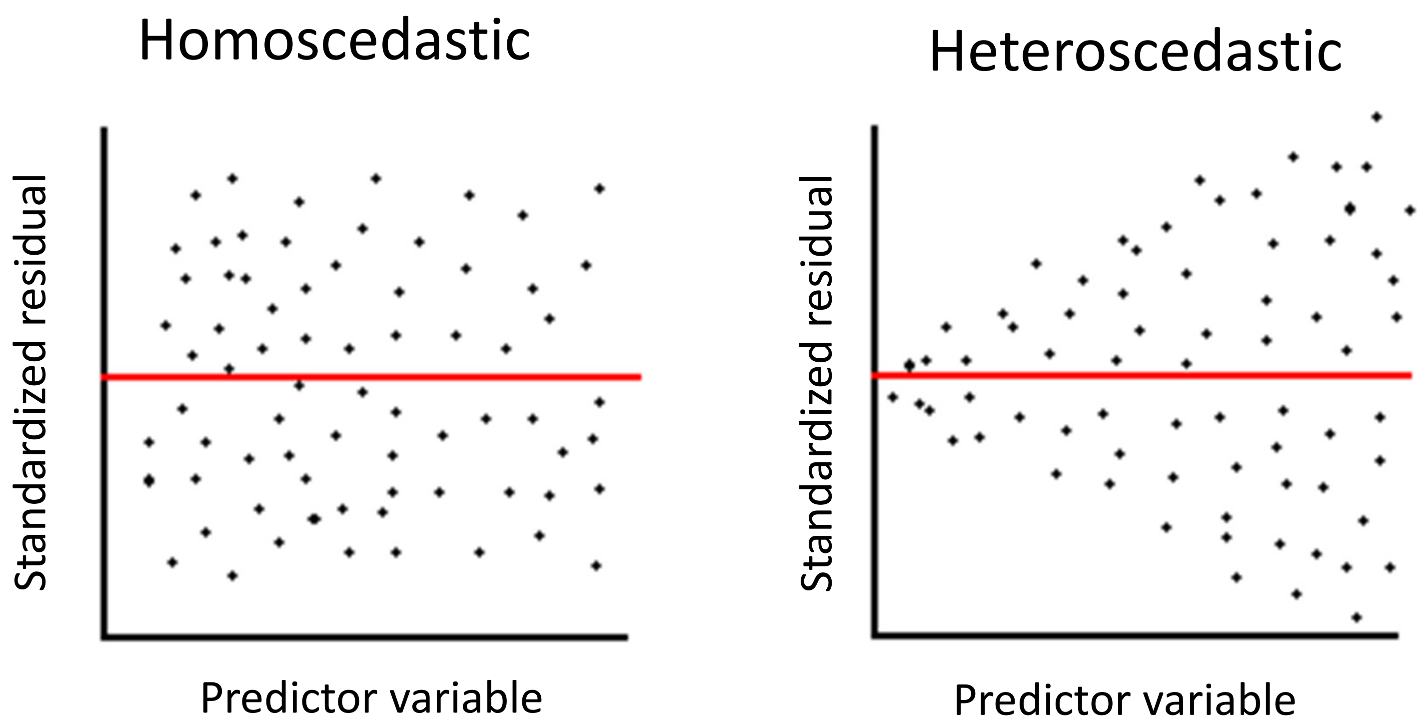
# ANOVA residuals



# Regression residuals



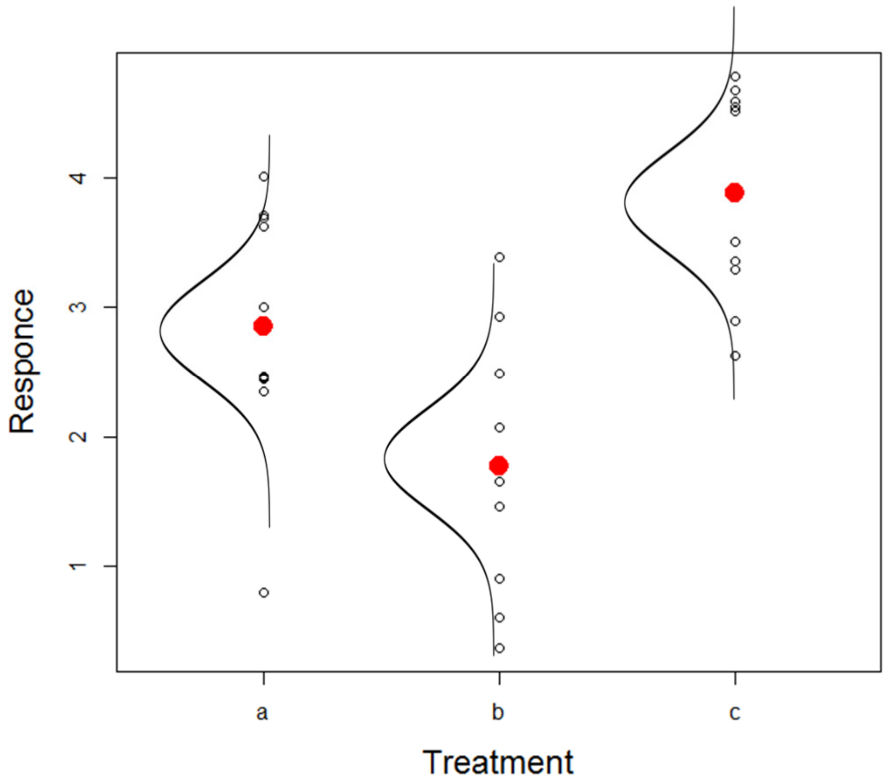
# Homoscedasticity



# Normality assumption

The assumption of normality applies to the variation around the expected value – the residual – not to the whole data set

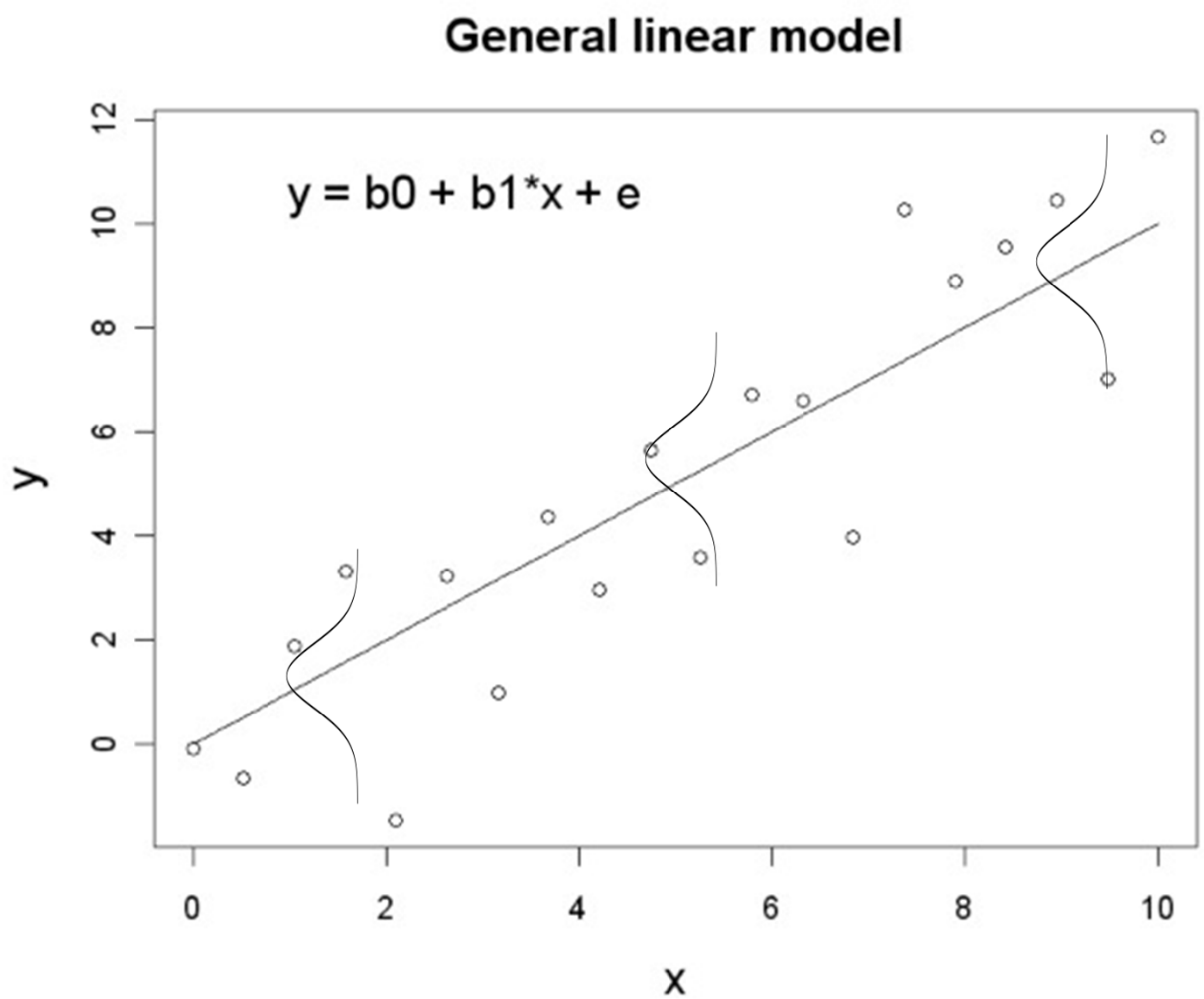
# Normality

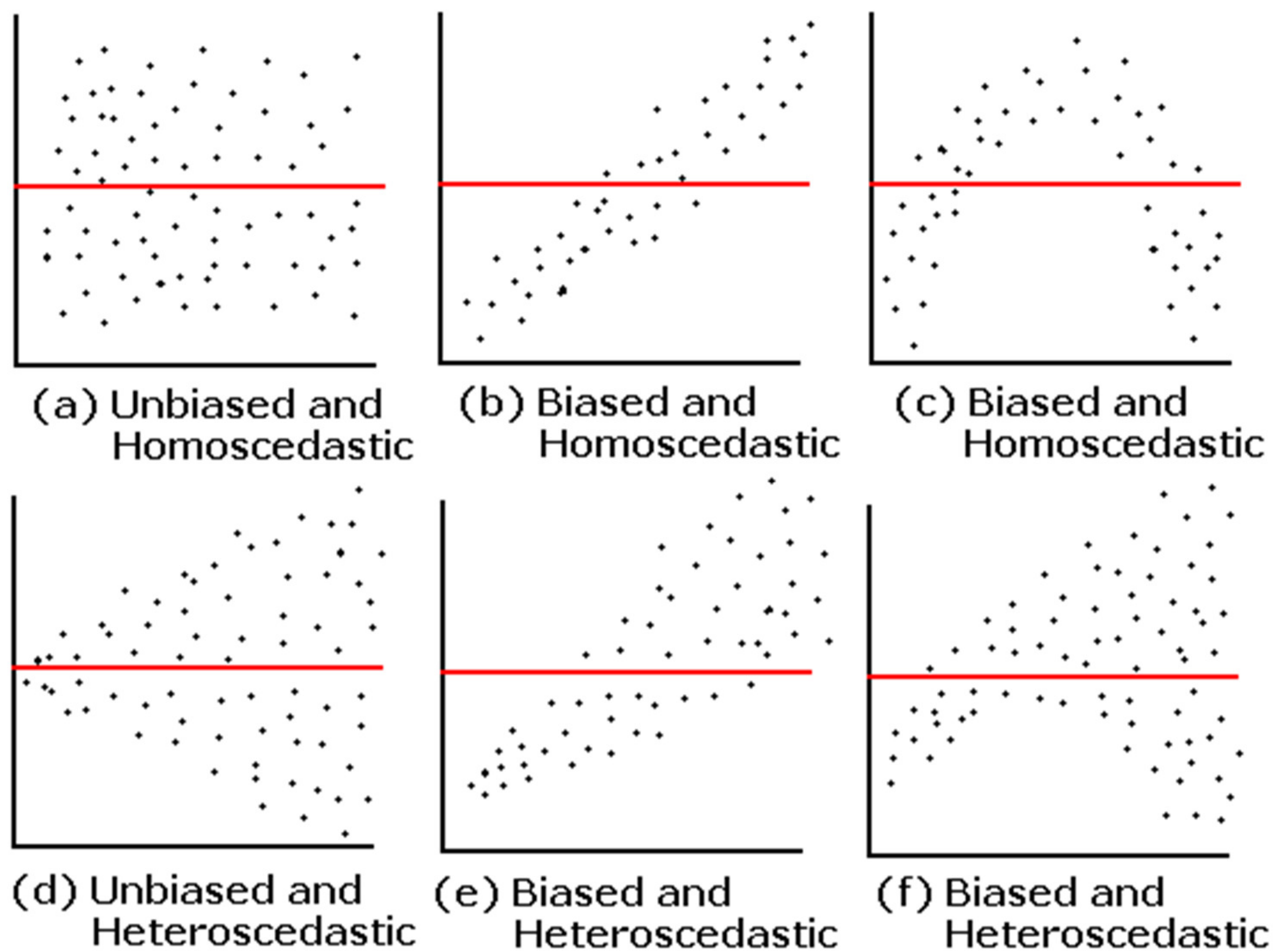



testing for ANOVA normality is easy

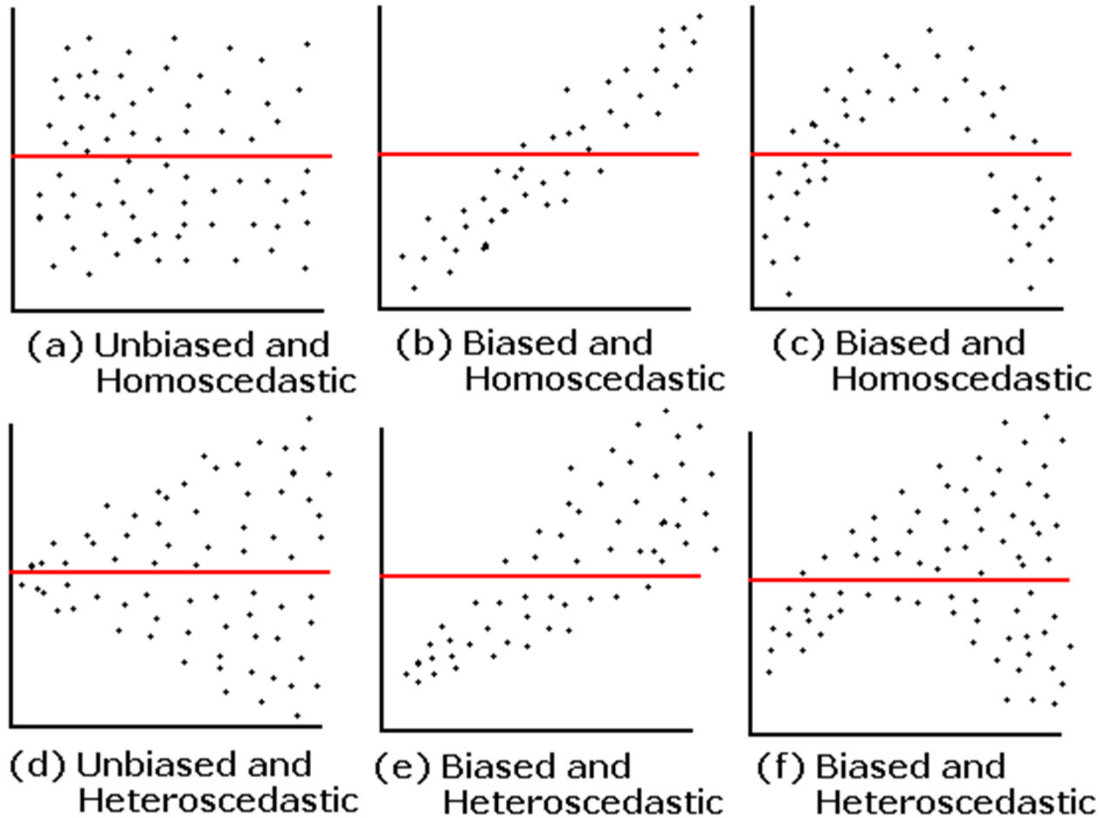
# Normality

$$Y \sim Normal(b_0 + \ b_1 X, \sigma^2)$$






What do these residuals tell us about the statistical model?




(a) Unbiased and homoscedastic. The residuals average to zero in each thin vertical strip and the SD is the same all across the plot.
(b) Biased and homoscedastic. The residuals show a linear pattern, probably due to a lurking variable not included in the experiment.
(c) Biased and homoscedastic. The residuals show a quadratic pattern, possibly because of a nonlinear relationship. Sometimes a variable transform will eliminate the bias.
(d) Unbiased, but homoscedastic. The SD is small to the left of the plot and large to the right: the residuals are heteroscadastic.
(e) Biased and heteroscedastic. The pattern is linear.
(f) Biased and heteroscedastic. The pattern is quadratic.

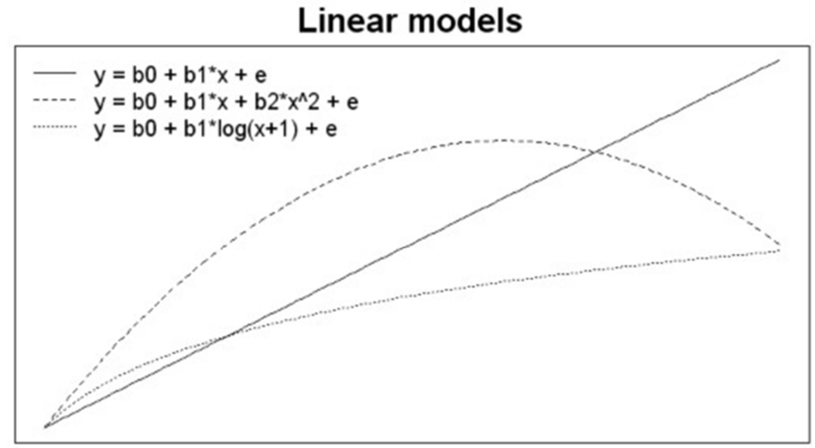
# Linear models

Example linear models:

$$Y \sim Normal\left(b_0 + b_1 x, \sigma^2\right)$$

$$Y \sim Normal\left(b_0 + b_1 x + b_2 x^2, \sigma^2\right)$$

$$Y \sim Normal\left(b_0 + b_1 \log(x), \sigma^2\right)$$




In all of these models we can define a new explanatory variable Z, such that the model can be written in the common linear equation form

$$Y \sim Normal\left(b_0 + b_1 z, \sigma^2\right)$$

Linear does not mean that the relationship between *Y* and *X* is linear; it simply means that *Y* can be expressed as a linear function of *X.*

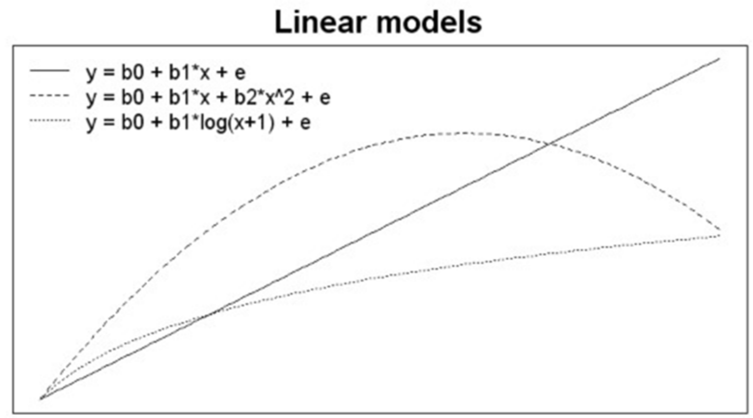
# Linear vs. non-linear models

Example linear models:

$$Y \sim Normal\left(b_0 + b_1 x, \sigma^2\right)$$

$$Y \sim Normal\left(b_0 + b_1 x + b_2 x^2, \sigma^2\right)$$

$$Y \sim Normal\left(b_0 + b_1 \log(x), \sigma^2\right)$$

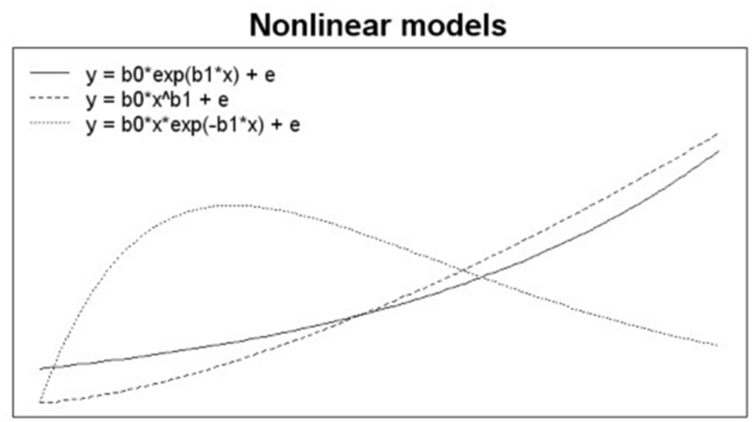



Example nonlinear models:

$$Y \sim Normal\left(b_0 e^{b_1 x}, \sigma^2\right)$$

$$Y \sim Normal\left(b_0 x^{b_1}, \sigma^2\right)$$

$$Y \sim Normal\left(b_0 x e^{-b_1 x}, \sigma^2\right)$$

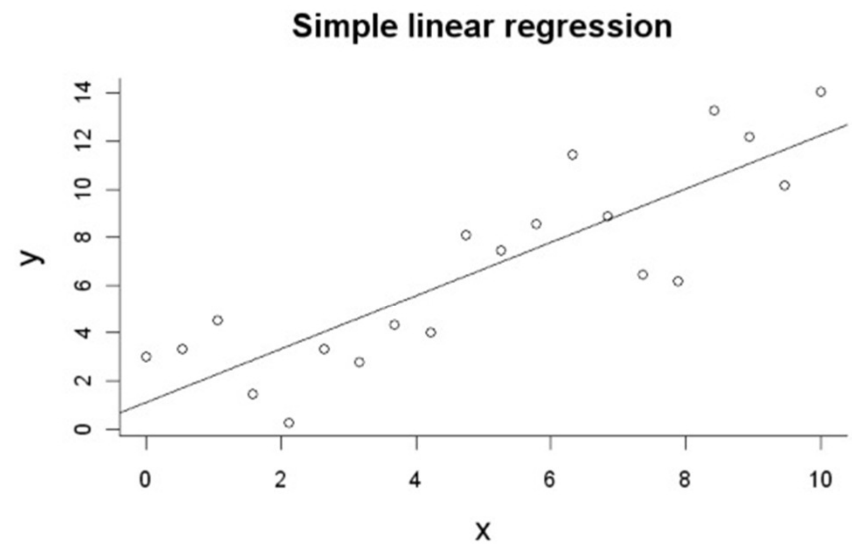



In these examples the models are all linear with respect to $b_0$ but nonlinear with respect to $b_1$

Simple linear regression

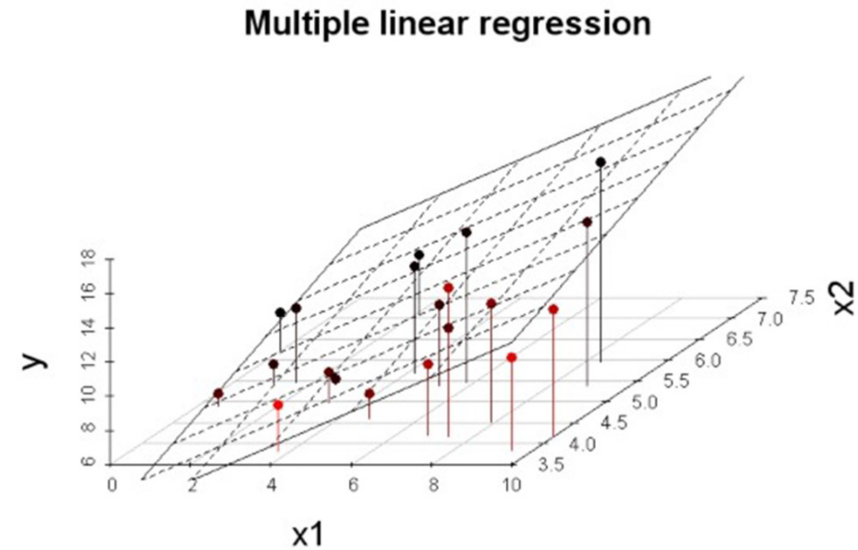
- Single continuous predictor

$$Y \sim Normal\left(b_0 + b_1 x, \sigma^2\right)$$



# Multiple linear regression

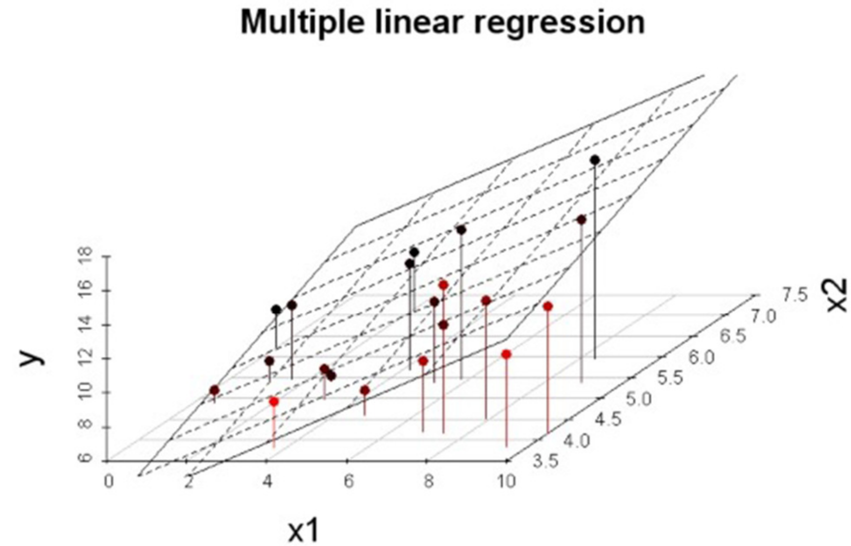
- Multiple continuous predictors




$$Y \sim Normal\left(b_0 + b_1 x_1 + b_2 x_2 + \ldots, \sigma^2\right)$$

## Multiple linear regression

- Multiple continuous predictors



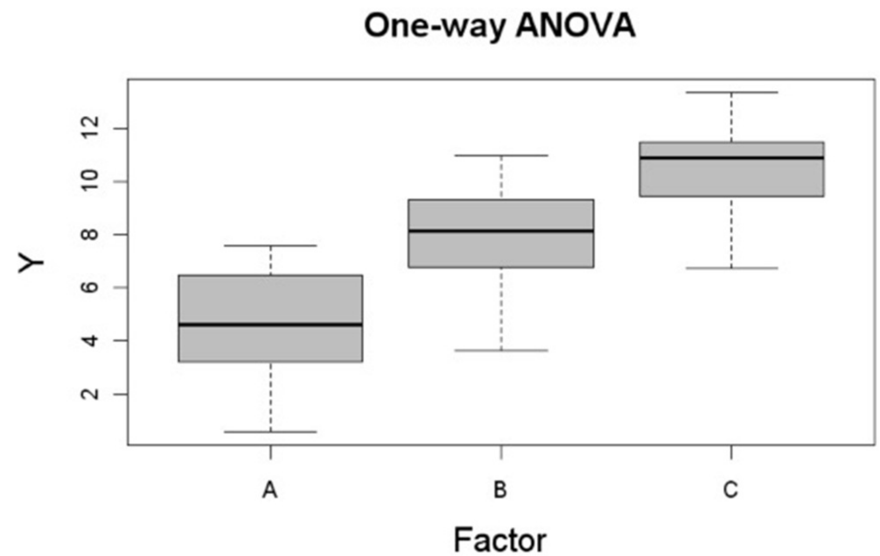

$$Y \sim Normal\left(b_0 + b_1 x_1 + b_2 x_2 + \ldots, \sigma^2\right)$$

In addition, interactions among covariates can be added. This tests whether the slope with respect to one covariate changes linearly as a function of another covariate.

$$Y \sim Normal\left(b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2, \sigma^2\right)$$

One-way analysis of variance (ANOVA)

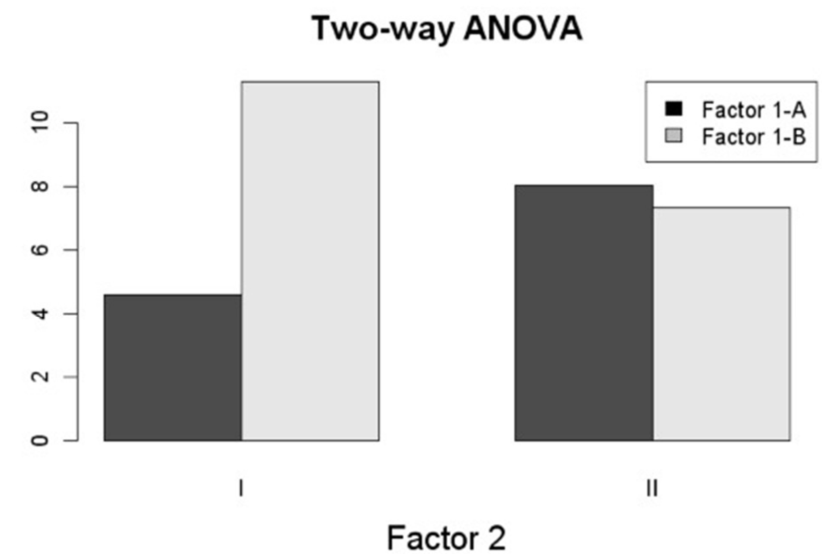
- Single categorical predictor (factor)




$$Y_i \sim Normal\left(\alpha_i, \sigma^2\right)$$

# Multiway ANOVA

- Multiple categorical predictors (factors)



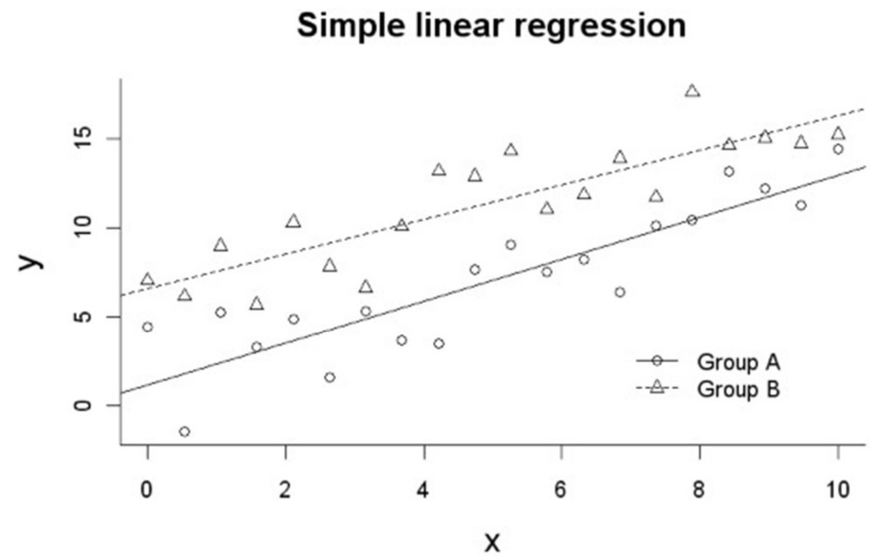

$$Y_{ij} \sim Normal\left(\alpha_i + \beta_j + \gamma_{ij}, \sigma^2\right)$$

Interaction

# Analysis of covariance (ANCOVA)

- Mix of categorical predictors (factors) and continuous covariate

$$Y_i \sim Normal\left(\alpha_i + \beta_i x, \sigma^2\right)$$



# General linear models

- Linear regression
- One- and multi-way analysis of variance (ANOVA)
- Analysis of covariance (ANCOVA)

- In R these procedures are use the function lm()